

Received April 17, 2020, accepted April 26, 2020, date of publication April 30, 2020, date of current version May 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991424

Residual Convolutional Neural Network for Cardiac Image Segmentation and Heart Disease Diagnosis

TAO LIU¹, YUN TIAN¹, SHIFENG ZHAO¹, XIAOYING HUANG¹, AND QINGJUN WANG²

¹College of Artificial Intelligence, Beijing Normal University, Beijing 100875, China

²Sixth Medical Center, Chinese PLA General Hospital, Beijing 100048, China

Corresponding author: Yun Tian (tianyun@bnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61472042 and Grant 61802020, in part by the Beijing Natural Science Foundation under Grant 4174094, and in part by the Fundamental Research Funds for the Central Universities under Grant 2015KJJC25.

ABSTRACT Deep learning (DL) has been widely used in biomedical image segmentation and automatic disease diagnosis, leading to state-of-the-art performance. However, automated cardiac disease diagnosis heavily relies on cardiac segmentation maps from cardiac magnetic resonance (CMR), most current DL segmentation methods, such as 2D convolution on planes, 3D convolution, are not fully applicable to CMR due to loss of spatial structure information or large gap between slices. To make better exploit spatial aspects of the CMR data to improve cardiac segmentation accuracy, we propose a new DL segmentation structure, which consists of a residual convolution neural network for compressing the intra-slice information, and a bidirectional-convolutional long short term memory (Bi-CLSTM) for leveraging the inter-slice contexts. Moreover, automatic disease diagnosis has been conducted using the segmentation maps. Experimental results of the automatic cardiac diagnosis challenge (ACDC) show that our cardiac segmentation structure and disease diagnosis methods have achieved promising results and it can be widely extended to computer-aided diagnosis.

INDEX TERMS Image segmentation, cardiac disease diagnosis, deep learning.

I. INTRODUCTION

Cardiac magnetic resonance (CMR) is the gold standard method for the assessment of cardiac structure and function [1], extracting left and right ventricular ejection fractions (EF), stroke volumes (SV) and left ventricle mass from CMR is usually as the primary step for cardiac disease diagnosis, but most current heart segmentation methods process 3D data in a slice-by-slice fashion, as the result these methods lose the spatial structure information in the original data, so we use the method of combining 2D and temporal information to compensate the spatial aspects loss of the original data in the 2D convolution network.

In this study, we propose a new DL segmentation structure, which is constituted by a residual convolution neural network for compressing the intra-slice information, and a Bi-CLSTM [2] for leveraging the inter-slice contexts.

The associate editor coordinating the review of this manuscript and approving it for publication was Inês Domingues¹.

Moreover, we conducted automatic disease diagnosis from the segmentation maps. Experimental results on the automatic cardiac diagnosis challenge show that our cardiac segmentation structure and disease diagnosis achieved promising results. In summary, our contributions are:

- The intra-slice information is compressed by convolutional neural network. U-Net is used as a submodule of our improved framework [3]. In order to mitigate the problem of vanishing/exploding gradient, we used the residual connection in our network structure [4]. At the same time, we used the dilated convolution instead of the max-pooling operation to improve the segmentation accuracy [5].
- Bi-CLSTM [2] is used to learn spatial information, which essentially assembles the intra-slice contexts into a 3D context and extracts 3D context information from it. For a set of 3D cardiac data, adjacent frames are similar, so the temporal relationship in Bi-CLSTM are used to simulate intra-slice contexts.

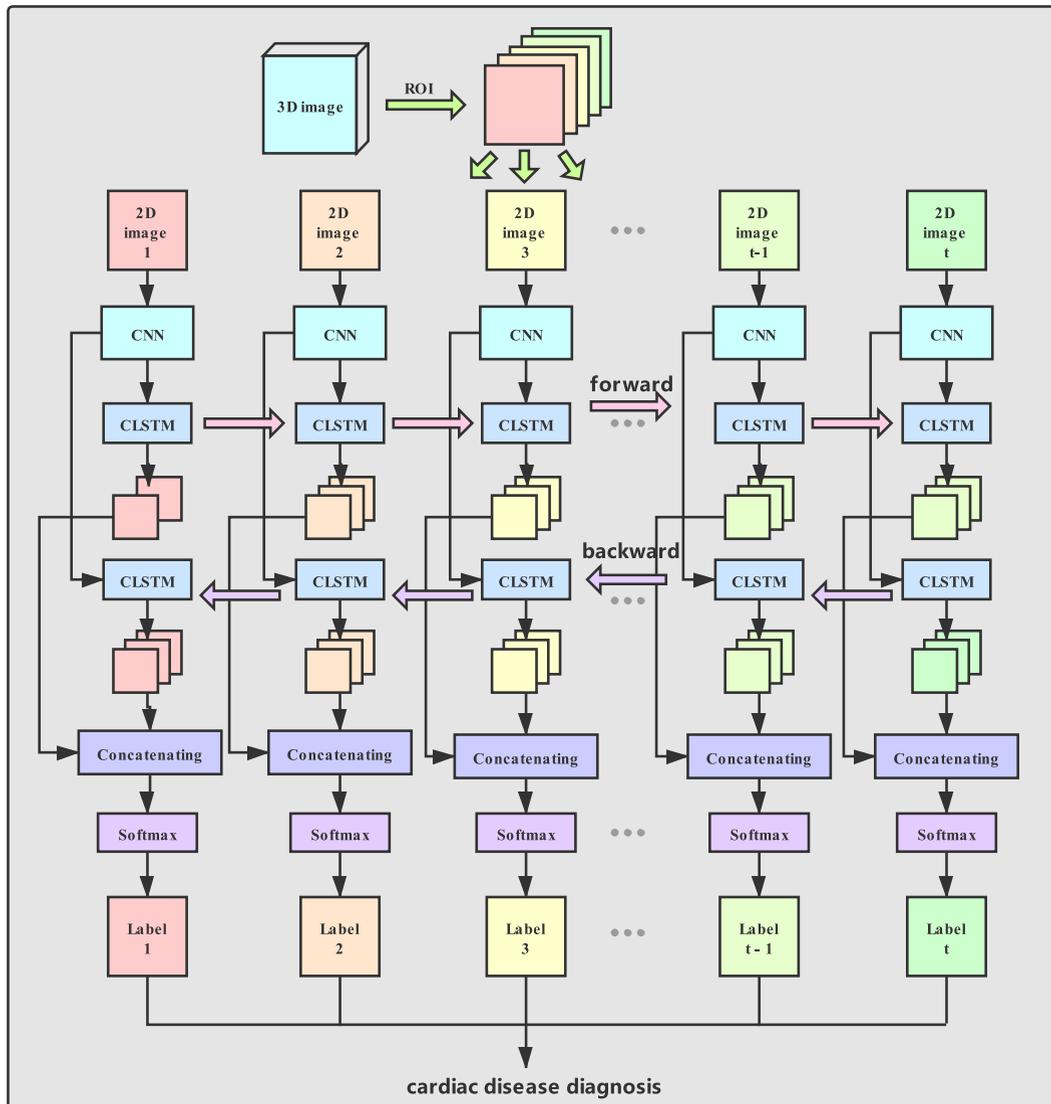


FIGURE 1. System overview of proposed method for automated cardiac segmentation and diagnosis. The pipeline involved: (i) The region of interest (ROI) is extracted from the CMR data, then, process the 3D data into 2D slices. (ii) The 2D slices are fed into the residual convolution neural network (CNN) (Fig. 2 shows the network structure of the residual convolution network in detail) for exploiting the intra-slice information. (iii) Input the results learned by residual convolution neural network into two CLSTMs (which work in two opposite directions, also known as Bi-CLSTM) for learning spatial information. (iv) Concatenate the results of the two CLSTMs learning, then a softmax function is applied to get the cardiac segmentation. (v) Finally, cardiac disease is automatically diagnosed from the segmentation maps.

- Cardiac diseases can be automatically diagnosis from the segmentation maps. Our approach combines automatic cardiac segmentation and disease diagnosis together. As shown in Fig. 1, which is a complete framework.

II. RELATED WORK

Each frame in the CMR is a 3D structure, current state-of-the-art methods for 3D biomedical image segmentation (Here, we mainly focus on DL segmentation methods, because they are relevant to our scheme and usually perform well) can be roughly divided into the following five categories. (I) 2D convolution on planes. (II) 3D convolution.

- (III) Multi-planar schemes. (IV) Fusion of 2D and 3D. (V) Combining 2D and temporal information.

A. 2D CONVOLUTION ON PLANES

U-Net only needs less training data and gets better segmentation results [3], so such network structure is considered as the most prominent 2D network structure in the field of medical image segmentation. Baumgartner *et al.* [6] used 2D U-Net to segment the CMR data, they processed the 3D data in a slice-by-slice fashion, and there is no correlation between each 2D slice, as the result this method lose the spatial structure information in the original data. In general, 2D medical image segmentation methods [7], [8] have the problem of losing spatial information.

B. 3D CONVOLUTION

The 3D convolution network structure is an improvement on 2D convolution, which makes full leverage of the spatial correlation along the z -direction. Yang *et al.* [9] used 3D convolution with residual connections for CMR image segmentation. However, due to large inter slice gap, the 3D convolutional network is not fully applicable to CMR image. Moreover, 3D convolution network usually has a complicated structure that requires many training parameters and a long training time. In the case of a small amount of training data, it tends to lead to over-fitting and poor generalization.

C. MULTI-PLANAR SCHEMES

Multi-planar method slices the dataset along different axial directions and compensates for the missing spatial structure information about the 2D segmentation method via multiple axial training methods. Prasoon *et al.* [10] integrated three 2D convolutional neural networks (xy , yz and zx planes of 3D image) to segment the tibial cartilage, and it performs better than the method based 3D multi-scale features. Similarly, Mortazi *et al.* [11] used multi-planar deep convolutional neural network to segment the whole heart and this method automatically utilizes spatial information by an adaptive fusion strategy. However, the multi-planar method is not suitable for CMR data, usually, CMR slice thickness from 5 mm to 10 mm, inter-slice gap of 5 mm, and the spatial resolution varies from 1.34 to 1.68 mm²/pixel [12], so the slices of xy plane have very different appearances in the slices of xz or yz plane, and slicing directly along different direction axes may cause unnecessary errors.

D. FUSION OF 2D AND 3D

At present, there are some 2D and 3D fusion network structures, and the segmentation accuracy is seriously relying on the fusion strategy. Isensee *et al.* [13] used a 2D and a 3D model for segmentation prediction and extracting features of averaged segmentation, which are fed into a classifier for cardiac disease diagnosis. This method of averaging 2D and 3D fusion models is simple but effective. Similarly, Li *et al.* [14] proposed a hybrid densely connected U-Net (H-DenseUNet) for liver and tumor segmentation, which fuse of 2D and 3D network structures. These methods make full use of the advantages of 2D convolution and 3D convolution, but the segmentation accuracy is easily affected by the fusion strategy.

E. COMBINING 2D AND TEMPORAL INFORMATION

The combination of 2D and temporal information compensates the loss of the original data spatial structure of the 2D convolution network by using a time storage network such as long short-term memory (LSTM). Zhang *et al.* [15] proposed a multi-level convolutional long short-term memory (ConvLSTM) model to the segmentation of left ventricle myocardium. LSTM is often used in time series data. e.g. [16] used LSTM to estimate regional wall thicknesses.

Tseng *et al.* [17] proposed a 3D biomedical segmentation method that joints sequence learning for 2D slices and jointly learn the multi-modalities. Simultaneously, Chen *et al.* [18] proposed a framework of 3D biomedical segmentation, which uses a recurrent neural network to exploit the inter-slice contexts. Inspired by these work [17], [18], we propose a new DL segmentation structure, which is constituted by a residual convolution neural network and a bidirectional-convolutional LSTM, and this network structure with 2D and temporal information has great potential.

In brief, there are several problems with the current methods of DL-based 3D biomedical segmentation. (i) 2D convolution on planes method treats the 3D data in a slice-by-slice fashion, thus losing the spatial structure information in the original data. (ii) 3D convolution network usually has a complicated structure, requires many training parameters, which requires a long training time. (iii) Multi-planar method performs well in some 3D biomedical segmentation, but this method is not suitable for CMR data (large inter-slice gap on the CMR). (iv) The fusion network structure and the 2D convolution method with temporal information have achieved good performance in CMR data segmentation, and we believe that these two methods have great potential in the future.

III. METHODOLOGY

Fig. 1 illustrates the framework for automated cardiac segmentation and diagnosis. The pipeline involved: (i) The region of interest (ROI) extraction method is firstly designed to perform data preprocessing operations (The detailed ROI extraction method can be found in the work of Ioffe and Szegedy [20]), then, process the 3D data into 2D slices (It should be noted here that the original data and the corresponding label are 3D data). (ii) The 2D slices are fed into the residual convolution neural network for exploiting the intra-slice information (Fig. 2 shows the network structure of residual convolution neural network in detail). (iii) Input the results learned by residual convolution neural network into two CLSTMs (which work in two opposite directions, also known as Bi-CLSTM) for learning spatial information. (iv) Concatenate the results of the two CLSTMs learning to get the cardiac segmentation results. (v) Finally, cardiac disease is automatically diagnosis from the segmentation maps. In Fig. 1, Bi-CLSTM is used to model temporal relationship between adjacent frames. The input is an image sequence $Image = \{Image_t | t = 1, 2, \dots, T\}$ across time frames t and the output is the predicted label map sequence $Label = \{Label_t | t = 1, 2, \dots, T\}$. The output of Bi-CLSTM is a pixel-wise feature map at each time point t . The time information between different slices is linked by the Bi-CLSTM. Compared to the standard LSTM which analyses one-dimensional signals, Bi-CLSTM can solve image sequence prediction problems and learn spatial structure information in image sequences. Section III-A introduces our proposed residual convolution neural network, Section III-B presents the Bi-CLSTM, and Section III-C describes the cardiac disease prediction in this study.

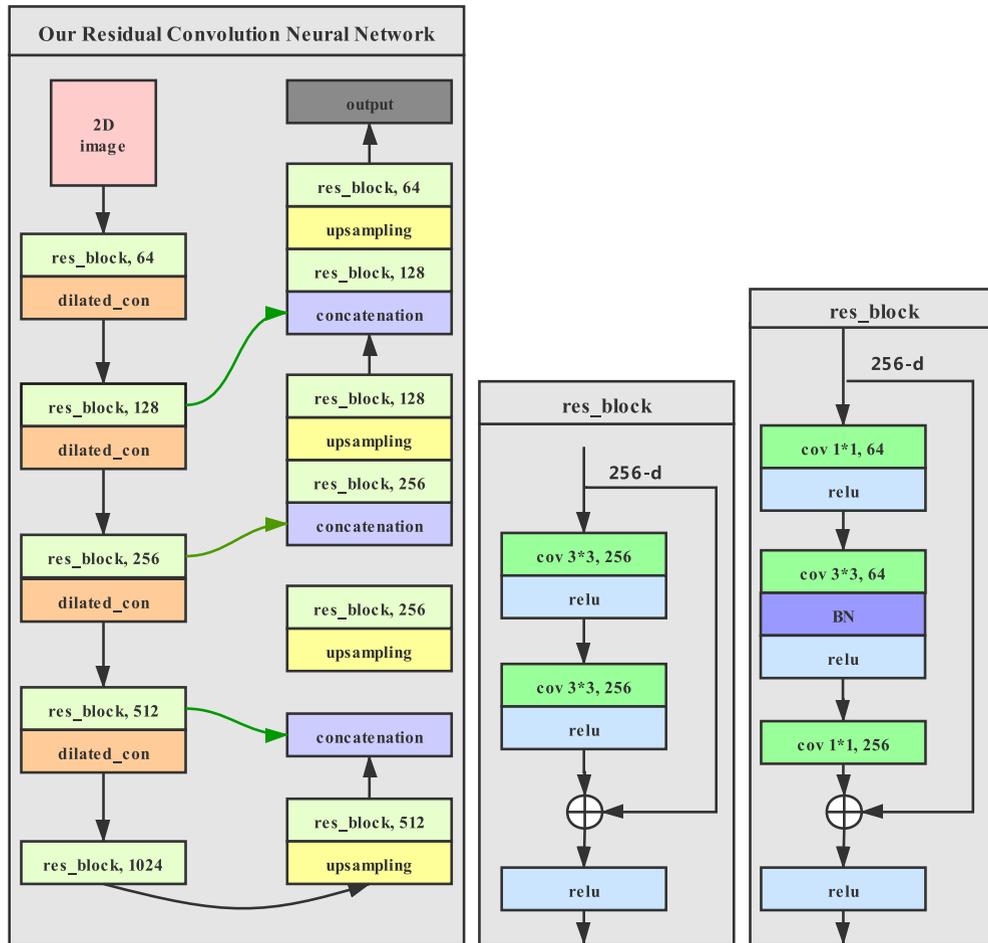


FIGURE 2. An overview of the residual convolution neural network and different residual network connections. (i) Left: the residual convolution neural network, note that the residual network is used to build deeper network structures, at the same time, the dilated convolution instead of the max-pooling operation. (ii) Middle: the normal residual connection, which uses a stack of 2 layers with a convolution kernel size of 3*3 and the feature layer of 256. (iii) Right: a deeper residual connection, which uses a stack of 3 layers. The first layer uses a 1*1 convolution kernel for dimensionality reduction, the “BN” in the second layer represents the batch normalization [19], and the third layer uses a 1*1 convolution for increasing dimensions. After the restoring and increasing dimensions of the first and third layers, the second layer has smaller input and output.

A. RESIDUAL CONVOLUTION NEURAL NETWORK

To obtain expressive and task-aware features (e.g., texture, shapes), a specially tailored residual convolution neural network is designed for each cardiac slice, as shown in the left of Fig. 2. Compared to the current convolutional neural network structure, our residual convolution neural network adds deep residual learning and dilated convolution.

Here, U-Net is employed as a submodule of our improved residual convolution neural network [3]. U-Net consists of a contracting path and a symmetric expanding path, which can capture context and enable precise localization. U-Net combines low-level features and high-level features, which is suitable for medical image segmentation. Therefore, U-Net is chosen as the network submodule, and the contracting path and the expanding path in U-Net are retained in our framework. Moreover, in our network structure, all blocks are replaced by residual connection, and the max-pool is replaced by dilated convolution.

1) DEEP RESIDUAL LEARNING

Deeper network structures can learn richer features, many of the visual tasks have benefit from deeper network structures [21], [22]. Simply increasing the number of stacked layers will cause gradient explosion/disappearance problems [23], which can tend to cause give rise to network degradation. Normalized initialization is often used to alleviate this problem [24]. Therefore, He *et al.* [4] proposed a deep residual learning method, which solved the gradient explosion/disappearance problems of deep network structure, and made the deeper network structure possible. Based on U-Net as a submodule, we apply the method of residual connection to our network structure, which can learn more information on each cardiac slice by increasing the depth of the network.

Residual connection approach (as shown on the middle side of Fig. 2) is widely used in deep networks, however, deeper network requires more parameters and computational

resources. In this study, in order to design a more efficient model, the normal residual connection is replaced by the deeper residual connection (as shown on the right side of Fig. 2). In the deeper residual connection, a $1 * 1$ convolution kernel is used in the first layer to dimension reduction, and another $1 * 1$ convolution is used in the last layer to increase dimensions. By this connection method, the network structure conforms to the input and output dimensions, and reduces training parameters.

2) DILATED CONVOLUTION

The max-pooling operation increases the receptive field by reducing the image size, which results in a loss of resolution, as a result the segmentation accuracy is decreased. Yu and Koltun [5] used dilated convolutions to systematically aggregate multi-scale contextual information without loss of resolution or coverage. Inspired by this, to improve the cardiac segmentation accuracy proposed in this study, we replaced all the max-pooling with the dilated convolution as shown on the left side of Fig. 2.

In brief, the details of our proposed residual convolution neural network structure as follows: (i) U-Net is used as the submodule. In order to capture context and enable precise localization, the contracting path and the expanding path structure are retained in our residual convolution neural network. (ii) In order to increase the depth of the network, learn more features and reduces training parameters, all blocks (a convolutional layer, a ReLU unit) in the U-Net are updated to a deeper residual connection (as shown on the right side of Fig. 2). (iii) All the max-pooling are replaced by the dilated convolution to increases the receptive field without loss of resolution. The residual convolution neural network structure is shown in the left side of Fig. 2.

B. BIDIRECTIONAL-CONVOLUTIONAL LONG SHORT TERM MEMORY: BI-CLSTM

Long short-term memory (LSTM) is a special recurrent neural networks (RNN) [25], LSTM and several variant units are widely used in processing sequential data [26]. LSTM can be able to perform better in longer sequences, and its definition

is as follows.

$$\begin{cases} i_t = \sigma(x_t W_{xi} + h_{t-1} W_{hi} + b_i) \\ f_t = \sigma(x_t W_{xf} + h_{t-1} W_{hf} + b_f) \\ c_t = c_{t-1} f_t + i_t \tanh(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \\ o_t = \sigma(x_t W_{xo} + h_{t-1} W_{ho} + b_o) h_t = o_t \tanh(c_t) \end{cases} \quad (1)$$

where i_t, f_t, c_t, o_t and h_t are respectively the input gate (i), forget gate (f), memory cell (c), output gate (o) and hidden state (h), at time t . W and b denote the diagonal weight matrices governing and bias, respectively. $\sigma()$ and $\tanh()$ are logistic sigmoid and hyperbolic tangent functions, x_t is the input at time t . There are many variants of LSTM, among them, Xingjian *et al.* [27] proposed a convolutional LSTM (CLSTM) network for the precipitation nowcasting problem, which replaces the vector multiplication by convolutional operators. The structure of CLSTM is shown in Fig. 3, where the left position enlarges its core computation unit. CLSTM can solve image sequence prediction problems and learn spatial structure information in image sequences. CLSTM can be formulated as follows.

$$\begin{cases} i_z = \sigma(x_z * W_{xi} + h_{z-1} * W_{hi} + b_i) \\ f_z = \sigma(x_z * W_{xf} + h_{z-1} * W_{hf} + b_f) \\ c_z = c_{z-1} \odot f_z + i_z \odot \tanh(x_z * W_{xc} + h_{z-1} * W_{hc} + b_c) \\ o_z = \sigma(x_z * W_{xo} + h_{z-1} * W_{ho} + b_o) h_z = o_z \odot \tanh(c_z) \end{cases} \quad (2)$$

Different from LSTM Eq. 1. $*$ denotes convolution and \odot denotes element-wise product. Bi-CLSTM [2] consists of two CLSTM in opposite directions, which allows the network model to learn more features. Concatenate the results of the two CLSTM learning, then a softmax function is applied to get the segmentation result.

The entire cardiac segmentation process is shown in Fig. 1. The pipeline involved: (i) Process the 3D data into 2D slices. (ii) Feed the 2D slices into the residual convolution neural network for exploiting the intra-slice information. (iii) Assembles the intra-slice contexts into 3D contexts, and input it into Bi-CLSTM for learning spatial information. (iv) Concatenate the results of the Bi-CLSTM learning to

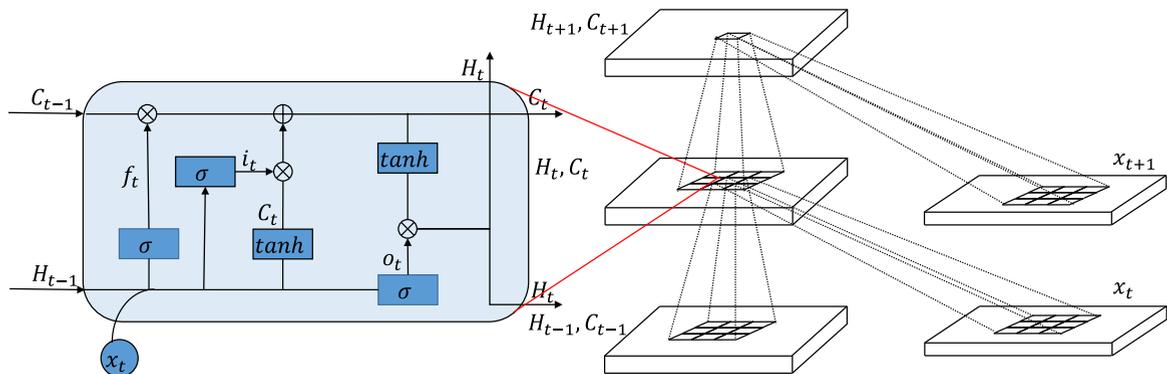


FIGURE 3. The structure of CLSTM.

get the cardiac segmentation. The left ventricular (LV), right ventricular (RV) and myocardium (Myo) for both end diastolic (ED) and end systolic (ES) phase instances can be segmented from the CMR by our proposed segmentation method.

C. CARDIAC DISEASE DIAGNOSIS

In this study, patients are divided into 5 classes according to the physiological parameters [12]: previous myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), abnormal right ventricle (RV) and without cardiac disease (NOR).

1) FEATURE EXTRACTION

Automatic cardiac disease diagnosis mainly uses patient characteristics and image features. The patient characteristics are patient weight (kg) and patient height (cm), and the image features (Left ventricular, right ventricular and myocardium for both end diastolic and end systolic phase instances) can be obtained from our segmented network structure. Combination of patient characteristics and image features can extract features such as ejection fraction (EF) and cardiac volume index. EF is commonly applied as a clinically relevant metric to assess ventricular function [28], and it can be calculated as follows.

$$EF = (EDV - ES) * 100\% / EDV \quad (3)$$

where, EDV and ES are ventricular end diastolic volume and ventricular systolic volume, respectively. The ejection fraction reflects the ejection function of the ventricle and is an important indication for determining the type of heart failure. The left ventricular volume index of diastolic end ($LVEDVI$) is calculated as follows:

$$LVEDVI = LVEDV / BSA \quad (4)$$

Here, $LVEDV$ is the left ventricular end diastolic volume, and BSA is the body surface area. The BSA can be calculated as $BSA = 0.007184 \cdot (weight^{0.425} \cdot height^{0.725})$ [29]. The extracted features are the basis for the cardiac disease diagnosis.

2) DIAGNOSTIC RULE

According to medical reports [12] that (i) patients with MINF have an ejection fraction of the left ventricle lower than 40%, (ii) patients with DCM have an ejection fraction of the left ventricle lower than 40% and the diastolic left ventricular volume $>100\text{mL}/\text{m}^2$, (iii) patients with HCM have a left ventricular cardiac mass high than $110\text{g}/\text{m}^2$, (iv) patients with RV have an ejection fraction of the right ventricle lower than 40% and the volume of the right ventricular cavity higher than $110\text{mL}/\text{m}^2$, (v) otherwise it is NOR. The cardiac disease diagnosis process is shown in Alg. 1 (It should be noted here that the parameters in Alg. 1 are derived from the medical report [12]).

Algorithm 1 Cardiac Disease Diagnosis Process

Require: Patient characteristics and image features

```

1: if  $EF_{LV} < 40$  or  $LVEDVI > 100$  then
2:   if  $LVEDVI > 100$  then
3:     Grouped into DCM
4:   else
5:     Grouped into MINF
6:   end if
7: else
8:   if  $(ED_{LV} + ED_{RV}) / ED_{Myo} < 2.4$  and  $ES_{Myo} + ED_{Myo} > 230$  then
9:     Grouped into HCM
10:  else
11:    if  $ES_{RV} > 110$  or  $EF_{RV} < 40$  then
12:      Grouped into RV
13:    else
14:      Grouped into NOR
15:    end if
16:  end if
17: end if

```

Ensure: Diagnostic result (one of MINF, DCM, HCM, RV, or NOR)

TABLE 1. Comparisons with different approaches on automated cardiac disease diagnosis.

Methods	Accuracy
Khened et al. [30]	0.96
Our method	0.94
Cetin et al. [31]	0.92
Isensee et al. [13]	0.92
Wolterink et al. [32]	0.86

IV. EXPERIMENT AND ANALYSIS

We have conducted experiments on the ACDC (Automated cardiac diagnosis challenge) datasets to evaluate the proposed method, and the ACDC took place during the MICCAI 2017 (International conference on medical image computing and computer assisted intervention), it remains open for new submissions over the next years. We compared the segmentation results with several state-of-the-art methods, and the segmentation results and diagnose results were evaluated on the dedicated online evaluation website.¹ We conducted experiments on a desktop computer with NVIDIA-GeForce-GTX-1080-Ti GPU, Intel Core i7-4790 CPU @ 3.60GHz and 32GB RAM.

A. SEGMENTATION RESULTS

We performed experiments on the ACDC dataset, which consisted of 150 patients (Patients with one of four diseases or healthy individuals), in addition, each patient has additional information: weight, height, ED, and ES. The ACDC dataset contains a training dataset and a testing dataset. (i) The training dataset composed of 100 patients including segmentation labels and diagnostic results, and our proposed network

¹<https://acdc.creatis.insa-lyon.fr/#challenges>

TABLE 2. Dice coefficients for cardiac segmentation. Note: *ED*: End diastole; *ES*: End systole; *LV*: Left ventricle; *RV*: Right ventricle; *Myo*: Myocardium.

Method	<i>ED</i>			<i>ES</i>		
	<i>LV</i>	<i>RV</i>	<i>Myo</i>	<i>LV</i>	<i>RV</i>	<i>Myo</i>
U-Net	0.95 (0.017)	0.89 (0.023)	0.84 (0.028)	0.87 (0.064)	0.82 (0.110)	0.86 (0.025)
U-Net+CLSTM	0.94 (0.016)	0.87 (0.026)	0.81 (0.031)	0.86 (0.046)	0.78 (0.121)	0.84 (0.026)
U-Net+Bi-CLSTM	0.94 (0.015)	0.90 (0.028)	0.84 (0.027)	0.86 (0.054)	0.84 (0.084)	0.84 (0.029)
Our RCNN	0.95 (0.017)	0.89 (0.030)	0.85 (0.027)	0.86 (0.043)	0.85 (0.094)	0.86 (0.034)
Our RCNN+Bi-CLSTM	0.95 (0.019)	0.92 (0.030)	0.86 (0.030)	0.87 (0.039)	0.86 (0.103)	0.87 (0.032)

TABLE 3. Hausdorff surface distance for cardiac segmentation. Note: *ED*: End diastole; *ES*: End systole; *LV*: Left ventricle; *RV*: Right ventricle; *Myo*: Myocardium.

Method	<i>ED</i>			<i>ES</i>		
	<i>LV</i>	<i>RV</i>	<i>Myo</i>	<i>LV</i>	<i>RV</i>	<i>Myo</i>
U-Net	19.18 (4.23)	29.46 (5.32)	28.24 (6.60)	26.99 (4.31)	40.46 (5.21)	32.81 (6.74)
U-Net+CLSTM	12.35 (4.01)	14.61 (5.24)	15.81 (6.01)	13.44 (4.11)	16.85 (4.20)	13.40 (4.23)
U-Net+Bi-CLSTM	14.34 (3.63)	17.61 (6.54)	18.81 (6.21)	17.24 (4.29)	26.75 (6.61)	18.40 (3.87)
Our RCNN	16.17 (4.73)	19.64 (7.39)	22.54 (6.94)	24.39 (5.91)	34.77 (6.33)	30.24 (4.43)
Our RCNN+Bi-CLSTM	11.91 (4.06)	17.90 (6.28)	19.18 (6.33)	15.78 (4.54)	24.72 (5.97)	20.69 (4.12)

model is trained on the training dataset. (ii) The testing dataset consisted of 50 new patients without segmentation labels and diagnostic results, we performed experiments on the testing dataset and evaluated our results on the dedicated online evaluation website.

The original input data size is not uniform, which is not conducive to subsequent processing. Therefore, the region of interest (ROI) is extracted from the original data. Gaussian kernel-based circular Hough transform approach was used for left ventricle (LV) localization, the original data is cropped according to the coordinates of the left ventricle center. After the ROI extraction, the 3D data is processed into 2D slices of $128 * 128$ size. The ROI extraction has the following three advantages: (i) Reducing interference from other tissues and organs. (ii) Unifying data format. The data slice is unified with $128 * 128$ size of convenient for network training. (iii) Saving computing resources.

It should be mentioned that the Dice coefficients and Hausdorff distance evaluation methods are used to evaluate the segmentation method in this study, and this is to prevent the segmentation algorithm from favoring some methods over others. Table 2 shows Dice coefficients for LV, RV and Myo segmentation at ED and ES. The Dice coefficients is to measure the accuracy of the segmentation results, and is defined as $D = 2(|V_{user} \cap V_{ref}|) / (|V_{user}| + |V_{ref}|)$, where V_{user} is the segmented volume and the V_{ref} is the corresponding reference volume. Dice coefficients value close to 1 represents a better segmentation result. Hausdorff distance measures the local maximum distance between the segmented surface S_{user} and the corresponding reference surface S_{ref} [33]. Table 3 shows the Hausdorff distance under different segmentation methods. The small Hausdorff distance represents a better segmentation result. Our method gets the Dice coefficients of the LV at ED of 0.95, the most difficult segmentation is the RV at ES of 0.86. Our approach has achieved satisfactory

results because the network can integrate contextual information along the z -direction.

In order to prove the effectiveness of the algorithm, we selected typical method for comparison. (i) U-Net method was selected to verify the effect of 2D segmentation method. (ii) U-Net + CLSTM and (iii) U-Net + Bi-CLSTM were used to prove the segmentation method of 2D combined temporal information. (iv) Our residual convolution neural network method was selected to verify the deep residual connection effect of 2D segmentation method. (v) Our method is a variant of 2D fusion temporal information approach. Comparing the U-Net and U-Net + Bi-CLSTM methods, the segmentation method with fusion temporal information has a smaller Hausdorff distance, that means the distance between the S_{ref} and the S_{user} is closer. The method proposed in this paper has a small improvement under the Dice coefficients (Table 2), but performs better under the Hausdorff distance evaluation index (Table 3), because there are certain differences in different evaluation indexes. Comparing the U-Net and our residual convolution neural network methods, residual convolution neural network has better segmentation results. Table 3 shows that Hausdorff distance of the U-Net+CLSTM method is the smallest except the LV Hausdorff distance of ED, which shows that the time information can improve the contour accuracy of segmentation. Our residual convolutional networks have deeper network structures to extract richer semantic features, and the connection of residuals makes the network learn different levels of semantic information. In addition, dilated convolution of the residual convolutional networks increases the receptive field of the network to improve segmentation accuracy. Moreover, it can be seen from Table 2 and Table 3 that the proposed method has a good performance under different evaluation metrics (Dice coefficients and Hausdorff distance), which proves that our method does combine intra-slice information and inter-slice

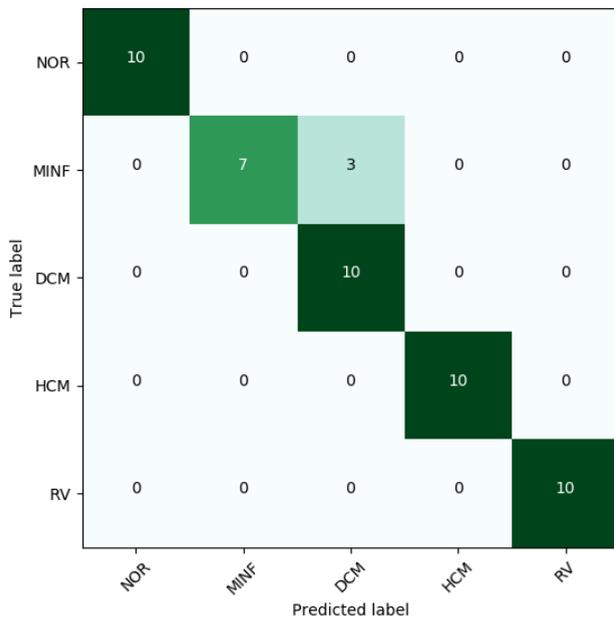


FIGURE 4. Classification metrics for automated cardiac disease diagnosis on ACDC challenge dataset (n = 50).

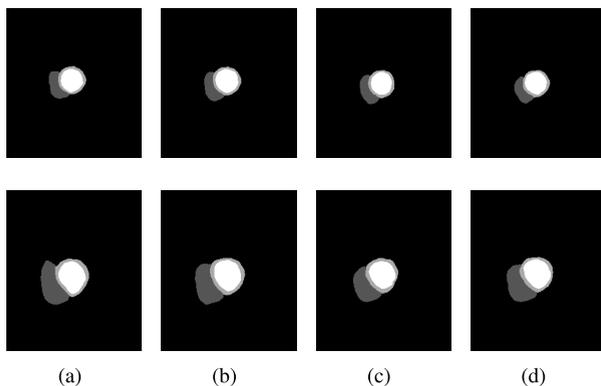


FIGURE 5. Visualization of cardiac segmentation results. (a): U-Net; (b) U-Net + Bi-CLSTM; (c) Our RCNN; (d) Our RCNN + Bi-CLSTM.

information from the sequenced image slices, so the segmentation accuracy is improved.

B. DIAGNOSE RESULTS

Cardiac disease diagnosis rule described in Section III-C. The diagnostic results obtained were evaluated on the online evaluation website. Fig. 4 shows the confusion matrix of our results on automated disease classification challenge, with an overall accuracy of 94%. Sensitivity was 100% for the NOR, DCM, HCM and RV classes, 70% for the MINF class. The three MINFs were mistakenly divided into DCM because the two diseases are visually similar and characterized by low EF of the LV. Table 1 compares our method with other cardiac disease diagnosis methods, and our method has a promising result.

V. CONCLUSION

In this study, we proposed a new DL segmentation structure, which is based on a residual convolution neural network

to compress the intra-slice information and a Bi-CLSTM to leverage the inter-slice contexts. Using this segmentation structures, the left ventricle, right ventricle, and myocardium were successfully segmented from the CMR dataset. In addition, we conducted automatic cardiac disease diagnosis from the segmentation maps. Our approach combines automatic cardiac segmentation and diagnosis, which has great potential for computer-aided diagnosis. We performed experiments on the public ACDC dataset, and the results of automated segmentation results and diagnostic results were satisfactory. In the future, we will delve deeper into the segmentation method for 2D combined temporal information, because this segmentation method has great potential in 3D biomedical segmentation.

REFERENCES

- [1] S. E. Petersen, N. Aung, M. M. Sanghvi, F. Zemrak, K. Fung, J. M. Paiva, J. M. Francis, M. Y. Khanji, E. Lukaschuk, A. M. Lee, V. Carapella, Y. J. Kim, P. Leeson, S. K. Piechnik, and S. Neubauer, "Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in caucasians from the UK biobank population cohort," *J. Cardiovascular Magn. Reson.*, vol. 19, no. 1, p. 18, Dec. 2017.
- [2] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 12, p. 1330, 2017.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [5] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: <http://arxiv.org/abs/1511.07122>
- [6] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, "An exploration of 2D and 3D deep learning techniques for cardiac mr image segmentation," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Springer, 2017, pp. 111–119.
- [7] M. Baldeon-Calisto and S. K. Lai-Yuen, "Adaesu-net: Multiobjective adaptive convolutional neural network for medical image segmentation," *Neurocomputing*, to be published.
- [8] H. Hu, N. Pan, J. Wang, T. Yin, and R. Ye, "Automatic segmentation of left ventricle from cardiac MRI via deep learning and region constrained dynamic programming," *Neurocomputing*, vol. 347, pp. 139–148, Jun. 2019.
- [9] X. Yang, C. Bian, L. Yu, D. Ni, and P.-A. Heng, "Class-balanced deep neural network for automatic ventricular structure segmentation," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart* Springer, 2017, pp. 152–160.
- [10] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2013, pp. 246–253.
- [11] A. Mortazi, J. Burt, and U. Bagci, "Multi-planar deep segmentation networks for cardiac substructures from MRI and CT," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Springer, 2017, pp. 199–206.
- [12] O. Bernard et al., "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018.
- [13] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Springer, 2017, pp. 120–129.
- [14] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

- [15] D. Zhang et al., "A multi-level convolutional LSTM model for the segmentation of left ventricle myocardium in infarcted porcine cine MR images," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Washington, DC, USA, 2018, pp. 470–473.
- [16] W. Xue, I. B. Nachum, S. Pandey, J. Warrington, S. Leung, and S. Li, "Direct estimation of regional wall thicknesses via residual recurrent neural network," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2017, pp. 505–516.
- [17] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, "Joint sequence learning and cross-modality convolution for 3D biomedical segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6393–6400.
- [18] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, "Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3036–3044.
- [19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [20] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers," *Med. Image Anal.*, vol. 51, pp. 21–45, Jan. 2019.
- [21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
- [23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
- [25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [26] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
- [27] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
- [28] P. L. M. Kerkhof, P. M. van de Ven, B. Yoo, R. A. Peace, G. R. Heyndrickx, and N. Handly, "Ejection fraction as related to basic components in the left and right ventricular volume domains," *Int. J. Cardiol.*, vol. 255, pp. 105–110, Mar. 2018.
- [29] D. Du Bois, "A formula to estimate the approximate surface area if height and weight be known," *Nutrition*, vol. 5, pp. 303–313, 1989.
- [30] M. Khened, V. Alex, and G. Krishnamurthi, "Densely connected fully convolutional network for short-axis cardiac cine mr image segmentation and heart diagnosis using random forest," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Springer, 2017, pp. 140–151.
- [31] I. Cetin, G. Sanroma, S. E. Petersen, S. Napel, O. Camara, M.-A. G. Ballester, and K. Lekadir, "A radiomics approach to computer-aided diagnosis with cardiac cine-MRI," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Springer, 2017, pp. 82–90.
- [32] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Automatic segmentation and disease classification using cardiac cine mr images," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Springer, 2017, pp. 101–110.
- [33] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Med. Image Anal.*, vol. 30, pp. 108–119, May 2016.



TAO LIU received the bachelor's degree in software engineering from Northwest A&F University, in 2018. He is currently pursuing the master's degree with the Institute of Virtual Reality and Visualization Technology, Beijing Normal University. His research interests include medical imaging and machine learning.



YUN TIAN received the B.Sc. degree in computer science and technology from Henan Normal University, in 2003, and the Ph.D. degree in signal and information processing from Northwestern Polytechnic University, China, in 2007. From 2014 to 2016, he was engaged in Postdoctoral research at the joint research station of the Chinese Academy of Engineering-Tsinghua University, China. He is currently an Associate Professor with the College of Information Science and Technology, Beijing Normal University. His research interests include medical image processing and pattern recognition.



SHIFENG ZHAO received the Ph.D. degree from Beijing Normal University, China, in 2014. She is currently a Lecturer with the Collage of Information Science and Technology, Beijing Normal University. Her research interests include medical imaging and virtual reality.



XIAOYING HUANG is currently studying at the School of Information Science and Technology, Beijing Normal University, Beijing, China, where she is a Student. Her research interest is in medical image processing.



QINGJUN WANG is currently the Deputy Director of medical imaging, a Medical Doctor, and the Deputy Chief Physician. Particularly good at post-treatment evaluation of brain tumors, preoperative white matter fiber bundle and central functional area localization analysis, cardiac magnetic resonance imaging, and cardiac function analysis.

...